

CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ

Rio de Janeiro - RJ - Brasil

ISSN 1413-9022 / v. 21: p. 01 - 10, 2006

UMA APLICAÇÃO DA MODELAGEM DE DEPENDÊNCIAS NA ANÁLISE DE RISCO DE PROJETOS

Ricardo de Melo e Silva Accioly
Instituto de Matemática e Estatística - UERJ
ricardo.accioly@gmail.com

Resumo

Na análise econômica de um projeto de desenvolvimento de um campo de petróleo são levados em consideração diversos elementos. Na parte de receitas o principal elemento é a curva de produção de petróleo. Os desembolsos podem ser divididos em dois grandes grupos: investimentos e custos operacionais. Os custos operacionais podem ser de forma geral divididos em três grandes grupos: custos da plataforma, custos de logística e custos de suporte em terra. Em maior ou menor grau cada um destes custos tem alguma dependência com relação ao nível de produção do sistema. Neste trabalho será apresentada uma técnica que permitirá a modelagem destas dependências através da construção de distribuições bivariadas. A técnica é baseada no conceito de função de dependência ou representação uniforme, no qual a função de dependência faz a ligação das distribuições marginais univariadas com sua representação multivariada. Estes resultados permitirão numa etapa posterior a ampliação da análise de risco destes projetos.

Palavras-chave: *Custo Operacional, Distribuições Bivariadas, Simulação de Monte Carlo.*

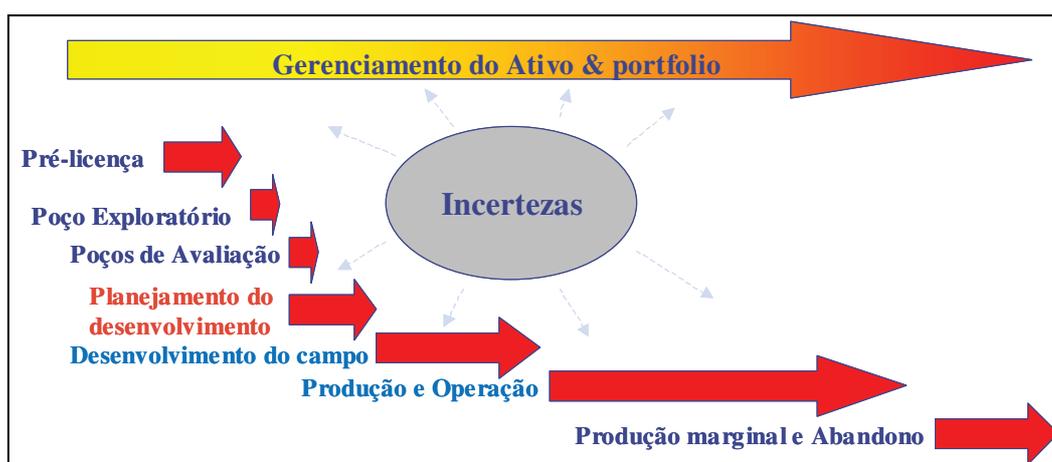
1. Introdução

Os empreendimentos na área de petróleo envolvem diversas vezes investimentos de bilhões de dólares, sendo que o retorno do capital investido é cercado de incertezas. As principais incertezas certamente são os preços do barril de petróleo durante toda a fase de produção e a curva de produção de petróleo. Na Bacia de Campos, localizada no litoral do Estado do Rio de Janeiro, a fase de produção, em geral, compreende cerca de vinte e cinco anos. Considerando que a etapa de planejamento e desenvolvimento de um campo está em torno de cinco anos, fica muito claro a importância de se ter uma modelagem adequada da fase de produção.

A representação esquemática do ciclo de vida de um projeto de Exploração e Produção pode ser visto na Figura 1. Entre a descoberta de um campo de petróleo até a decisão de desenvolvê-lo pode ter transcorrido diversos anos.

A análise econômica de um projeto de produção de um campo de petróleo é uma tarefa com elevado grau de complexidade. Existem diversos elementos envolvidos, cada um com grau de incertezas diferenciado, sendo que estas incertezas variam com a fase do empreendimento. É possível dividir de uma forma simplificada os componentes desta análise econômica. As receitas são basicamente relativas a produção de petróleo, os desembolsos podem ser divididos em dois grupos: investimentos de capital e custos operacionais.

Figura1: Ciclo de vida de um projeto de Exploração e Produção.



Fonte: Petrobras, 2006

Os investimentos são em sua maior parte relacionados à construção da plataforma de produção e dos poços de petróleo. Eles ocorrem na fase inicial, ou seja, na fase de desenvolvimento do campo. Os custos operacionais, que são o foco deste trabalho, ocorrem durante toda a fase de produção correspondendo a maior faixa de tempo a ser modelada na análise econômica.

Os custos operacionais podem ser divididos em três grandes grupos:

- Custos da plataforma;
- Custos de logística;
- Custo do suporte de terra

Estes custos em maior ou menor grau estão ligados aos níveis de produção de cada sistema. É muito comum a modelagem destes custos através de modelos lineares, pois isto simplifica muito a análise econômica. O uso de modelos lineares, ou mesmo linearizados através de transformações de variáveis, envolve suposições a respeito do comportamento das distribuições das variáveis envolvidas.

Seja o modelo de regressão linear simples dado por $y = \beta_0 + \beta_1 x + \varepsilon$, é conhecimento comum que certas suposições básicas são inerentes a este tipo de modelo, tais como:

- O termo correspondendo ao erro, ε , é uma variável aleatória normalmente distribuída. A variável resposta, y , é a soma de uma componente constante e uma variável aleatória normal, logo sua distribuição também é normal ao longo da regressão em valores de x .
- O erro tem valor esperado zero e variância σ_ε^2 . Isto implica que a distribuição dos y ao longo da regressão tem sempre a mesma variância, independente do valor de x .

Estas duas suposições dificilmente são respeitadas quando estamos analisando dados reais. No caso dos custos operacionais de plataformas, estes têm, em geral, comportamento bastante assimétrico e sua variação aumenta à medida que aumentam os volumes produzidos. Através de transformações de variáveis poderíamos corrigir parcialmente ou totalmente este problema. No entanto ao utilizar estas transformações estamos afetando diretamente a correlação entre as duas variáveis como podemos ver

em Dunlap *et al.* (1995). Ao alterar as correlações entre as variáveis indiretamente podemos estar introduzindo tendências não esperadas nos resultados.

Neste trabalho será apresentada uma solução diferente para modelagem da dependência entre duas variáveis. A construção de modelos multivariados é uma área de pesquisa que vem tomando impulso nos últimos anos com a crescente utilização de uma teoria que teve seu desenvolvimento inicial a partir da década de 40, conforme pode ser visto em Nelsen (1999).

2. Funções de dependência

O conceito da função de dependência pode ser compreendido pela representação dada por Nelsen (1999). Seja um par de variáveis aleatórias X e Y , com funções distribuições de probabilidade acumuladas $F(x) = P[X \leq x]$ e $G(y) = P[Y \leq y]$ respectivamente. Seja $H(x,y) = P[X \leq x, Y \leq y]$ a função de distribuição de probabilidade acumulada conjunta destas duas variáveis. É possível dizer que para cada par (x, y) temos associados três valores que estão no intervalo $[0, 1]$: $F(x)$, $G(y)$ e $H(x,y)$. Isto significa que cada par (x, y) de números reais está vinculado a um ponto $(F(x), G(y))$ está situado num quadrado unitário $[0, 1] \times [0, 1]$, sendo que este par ordenado corresponde a um número $H(x, y)$ no intervalo $[0, 1]$. Será mostrado a seguir que esta correspondência que atribui um valor à função conjunta para cada par ordenado, é feita através de uma função, denominada de função copula. O teorema fundamental que originou o estudo das copulas é (Sklar, 1959):

Seja H uma distribuição de probabilidade conjunta com distribuições marginais F e G . Então existe uma função C tal que para todo x, y em $[-\infty, +\infty]$,

$$H(x, y) = C(F(x), G(y))$$

Se F e G são contínuas então C é única; senão, C é unicamente determinada na Amplitude $F \times$ Amplitude G .

Uma interpretação importante a ser tirada do teorema de Sklar (1959) é que o problema de se obter uma distribuição bivariada (ou multivariada) foi separado em duas partes. A função de dependência se tornou independente das distribuições marginais, sendo que a função é que contem toda a estrutura de dependência da distribuição

conjunta. Outro ponto importante é que as distribuições marginais podem ser definidas de maneira independente não sendo necessário, por exemplo, impor que sejam idênticas.

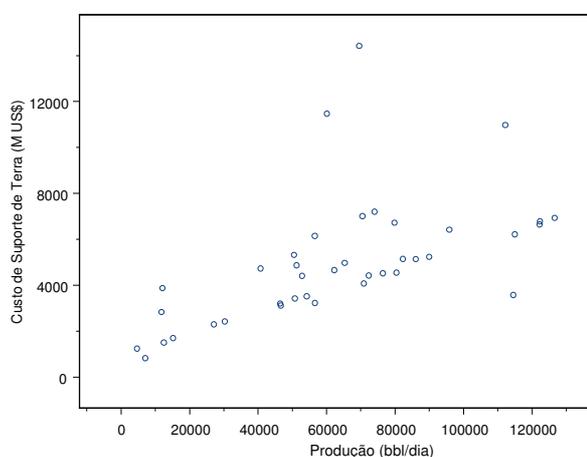
A medida de dependência mais frequentemente usada é o coeficiente de correlação de Pearson que é intrinsecamente ligado com a dependência linear e a distribuição normal. Quando estamos analisando dados cuja distribuição é normal, as médias e a variância-covariância podem ser modeladas separadamente e a dependência é caracterizada unicamente pelo coeficiente de correlação. Para outros tipos de dados, em geral, isto não é válido. A abordagem através das copulas representa a flexibilidade necessária quando modelando dados não-normais.

3. Estudo de caso

Vamos modelar a dependência entre o custo de suporte de terra e o volume de produção de um sistema. O custo de suporte de terra compreende todos os custos do ativo de produção devido ao pessoal técnico e administrativo, além dos custos corporativos tais como: recursos humanos, saúde, treinamento entre outros. Estes custos serão usados em conjunto com os ganhos da produção gerando um VPL simplificado do projeto.

A amostra utilizada para modelar os custos foi obtida de plataformas da Bacia Campos, ela continha trinta e nove custos de suporte de terra e as respectivas produções da plataforma. Para comparar os resultados obtidos a partir da construção de uma distribuição bivariada, foi também construído um modelo de regressão linear simples entre os custos e as produções.

Figura 2: Gráfico de dispersão da Produção da Plataforma e os Custos de Suporte de Terra



É visível a existência de um padrão semelhante de comportamento entre os níveis de produção e cada um dos custos, sendo isto mais evidente nos relacionamentos com custos de logística e custo de suporte em terra. A Tabela 1 apresenta os coeficientes de correlação de Pearson e o coeficiente de concordância de Kendall para a relação entre a produção de óleo e as demais variáveis.

Tabela 1: Comparação dos coeficientes de correlação de Pearson e concordância de Kendall.

	Produção (bbl/dia) x Custo de Suporte (M US\$)
Coef. de Pearson	0,566
Coef. de Kendall	0,557

Podemos perceber que o coeficiente de Kendall está indicando dependências levemente mais fracas que o coeficiente de Pearson. Entretanto o coeficiente de Kendall não é afetado por transformações das variáveis. Se for utilizada uma transformação logarítmica o valor coeficiente de Pearson passa a ser 0,706, enquanto que o de Kendall permanece inalterado.

Na Tabela 2 são apresentadas algumas de funções de dependência, seguindo a numeração de Nelsen (1999), que serão usadas para selecionar o melhor modelo. Para maiores referências sobre cada uma destas famílias consultar Nelsen (1999) páginas 94-97. Outra fonte de referência para estas e outras funções pode ser obtida no capítulo 5 de Joe (1997).

Para obter a estimativa dos parâmetros de cada uma das famílias foi usado o método apresentado por Genest *et al.* (1995). O primeiro passo foi à obtenção de uma função distribuição acumulada empírica aproximada através do método das B-Splines (KOOPEBERG & STONE, 1992). Na fase seguinte foram utilizados algoritmos de otimização, em um ambiente de programação de um pacote estatístico, para obter as estimativas dos parâmetros. Os resultados são apresentados na Tabela 3 em conjunto com as diferenças de AIC e os pesos de Akaike (BURNHAM & ANDERSON, 2002).

Tabela 2: Famílias de funções de dependência Arquimedianas com um parâmetro de dependência

Número/Nome	$C_\alpha(u, v)$	$\varphi_\alpha(t)$	$\alpha \in$
N.1 - Clayton	$(u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$	$\frac{1}{\alpha}(t^{-\alpha} - 1)$	$[-1, +\infty] \setminus \{0\}$
N.4 - Gumbel	$\exp(-[(-\ln u)^\alpha + (-\ln v)^\alpha]^{1/\alpha})$	$(-\ln t)^\alpha$	$[1, +\infty]$
N.5 - Frank	$-\frac{1}{\alpha} \ln \left(1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1} \right)$	$-\ln \frac{e^\alpha - 1}{e^{-\alpha} - 1}$	$(-\infty, +\infty) \setminus \{0\}$
N.13	$\exp(1 - [(1 - \ln u)^\alpha + (1 - \ln v)^\alpha - 1]^{1/\alpha})$	$(1 - \ln t)^\alpha - 1$	$(0, +\infty)$
N.14	$(1 + [(u^{-1/\alpha} - 1)^\alpha + (v^{-1/\alpha} - 1)^\alpha]^{1/\alpha})^{-\alpha}$	$(t^{-1/\alpha} - t)^\alpha$	$[1, +\infty)$

Tabela 3: Estimativas de MV do parâmetro de dependência, diferenças de AIC e pesos de Akaike.

Modelo	$\hat{\alpha}$	$L(\alpha)$	Δ_i	w_i
Clayton	2,293	20,034	0	0,53
N.13	5,461	19,025	1,01	0,32
N.14	1,345	16,595	3,44	0,09
Frank	6,758	15,694	4,34	0,06
N.6	1,864	5,154	14,88	0,00
Gumbel	1,300	3,808	16,23	0,00

Escolhendo o modelo de Clayton que apresentou o maior peso de Akaike foram simulados custos de suporte de terra para as produções da Tabela 4. Os custos gerados foram trazidos a valor presente e comparados com os resultados obtidos através da regressão linear.

Tabela 4: Dados de produção para simulação.

Ano	Vazão média esperada (bbl/dia)	Limite inferior da vazão média (bbl/dia)	Limite superior da vazão média (bbl/dia)
Ano 1	0	0	0
Ano 2	0	0	0
Ano 3	0	0	0
Ano 4	5.283,36	4.755,03	5.811,70
Ano 5	28.949,60	26.054,64	31.844,56
Ano 6	83.676,23	75.308,60	92.043,85
Ano 7	110.900,26	99.810,24	121.990,29
Ano 8	123.222,52	110.900,26	135.544,77
Ano 9	103.691,45	93.322,30	114.060,59
Ano 10	83.676,23	75.308,60	92.043,85

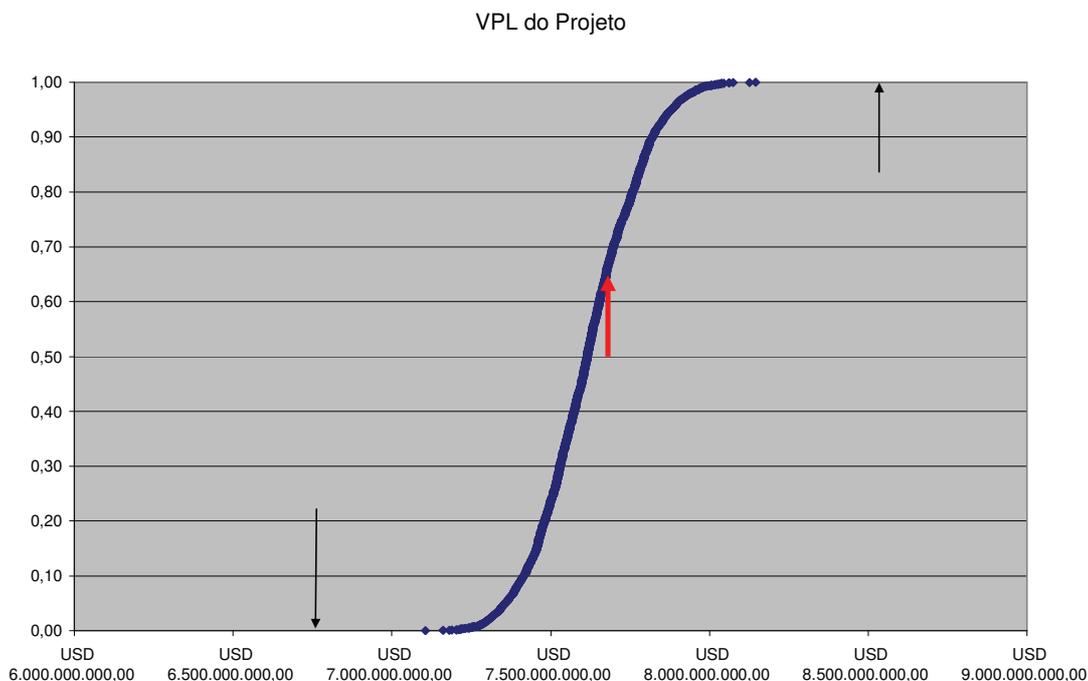
Ano 11	71.740,88	64.566,79	78.914,97
Ano 12	66.188,05	59.569,25	72.806,86
Ano 13	57.984,91	52.186,42	63.783,40
Ano 14	49.129,56	44.216,60	54.042,52
Ano 15	42.545,91	38.291,32	46.800,50
Ano 16	37.242,14	33.517,92	40.966,35
Ano 17	32.844,65	29.560,19	36.129,12
Ano 18	29.266,67	26.340,00	32.193,33
Ano 19	25.961,64	23.365,47	28.557,80
Ano 20	23.727,67	21.354,91	26.100,44
Ano 21	21.182,39	19.064,15	23.300,63
Ano 22	19.364,15	17.427,74	21.300,57
Ano 23	18.001,26	16.201,13	19.801,38
Ano 24	15.353,46	13.818,11	16.888,81
Ano 25	14.157,23	12.741,51	15.572,96
Ano 26	12.707,55	11.436,79	13.978,30
Ano 27	12.110,69	10.899,62	13.321,76

A regressão entre os custos de suporte de terra e os volumes de produção está representada no modelo linear abaixo:

$$\text{Custo de Suporte de Terra (US\$000)} = 2085,70 + 0,0460 \times \text{vazão de óleo (bbl/dia)}$$

A partir deste modelo linear forma obtidos os custos de suporte de terra para cada ano de produção.

Figura 3 Gráfico da distribuição acumulada do VPL do projeto



A Figura 3 apresenta os resultados da simulação da distribuição do VPL com a função de dependência de Clayton e também com a representação dos resultados da regressão (setas). Pela representação fica claro que a análise de sensibilidade obtida através do modelo de regressão tem pouco significado. O limite superior (US\$ 8.578.021.632,00) e o inferior (US\$ 6.783.264.435,00) estão fora da distribuição de VPL obtido pelo modelo bivariado. O valor obtido através da curva de produção esperada, US\$ 7.680.643.034,00, está dentro da distribuição estando num valor de probabilidade de 67%.

5. Conclusões

Neste trabalho foi apresentada uma técnica que permite a construção de distribuições bivariadas, que podem ser muito úteis em estudos de simulação. A função de dependência representa uma alternativa viável de modelagem de variáveis aleatórias não normais, permitindo obter modelos adequados aos padrões de dependência existentes entre as variáveis.

Os ganhos potenciais de seu uso em análises econômicas de projetos de Exploração e Produção ainda precisam ser avaliados num modelo completo, buscando a comparação com os procedimentos usuais.

Referências

- Burnham, K. P., Anderson, D. R., **Model Selection and Multimodel Inference – A Practical Information-Theoretic Approach**. 2ed. Springer, New York, 2002.
- Dunlap, W. P., Burke, M. J., Greer, T., The Effect of Skew on the Magnitude of Product-Moment Correlations, **The Journal of General Psychology**, 122(4), 365-377, 1995.
- Genest, C., Ghoudi, K., Rivest, Louis-Paul, A Semiparametric Estimation procedure of Dependence Parameters in Multivariate Families of Distributions. **Biometrika**, Volume 82, Issue 3, pp. 543-552, 1995.
- Joe, H. Multivariate Models and Dependence Concepts, **Monographs on Statistics and Applied Probability 73**, Chapman & Hall/CRC, 1997.
- Koopeberg, C., Stone, C., Log spline Density Estimation, **Journal of Computational and Graphical Statistics**, 1, 301-328, 1992.
- Nelsen, R. B., “An Introduction to Copulas”, **Lecture Notes in Statistics 139**, Springer, 1999.
- Sklar, A, “Fonctions de reparticion à n dimensions et leurs marges,” Publ. Inst. Statist. Unive. Paris 8, 229-231, 1959.

AN APPLICATION OF DEPENDENCY MODEL IN PROJECT RISK ANALYSIS

Abstract

When analyzing economic aspects of an upstream development project, some drivers are taken into account. Related with the revenues aspects, the main point is the oil production curve. Related with expenses we have two major groups: investments and operational costs. Operation costs, in general, can be classified in three major subgroups: platform costs, logistics costs and onshore support costs. All these elements have some kind of dependence related to the system production level. This article presents a technique that allows modeling these dependencies through the construction of bivariate distributions. This technique is based on the dependency function or uniform representation, in which the dependence function links the univariate margins with their full multivariate distribution. These results will allow, in a next step, the enhancement of risk analysis for this sort of projects.

Key-words: *Operational Cost, Bivariate Distribution, Monte Carlo Simulation.*